Data Value Networks: Enabling a New Data Ecosystem

Judie Attard, Fabrizio Orlandi, Sören Auer
University of Bonn
Germany
attard@iai.uni-bonn.de, orlandi@iai.uni-bonn.de, auer@cs.uni-bonn.de

***Abstract*—With the increasing permeation of data into all dimensions of our information society, data is progressively becoming the basis for many products and services. It is hence becoming more and more vital to identify the means and methods how to exploit the value of this data. In this paper we provide our definition of the *Data Value Network*, where we specifically cater for *non-tangible data products*. We also propose a *Demand and Supply Distribution Model* with the aim of providing insight on how an entity can participate in the global data market by producing a data product, as well as a concrete implementation through the *Demand and Supply as a Service*. Through our contributions we project our vision of generating a new *Economic Data Ecosystem* that has the Web of Data as its core.**

***Keywords*-value creation, data value network, data value chain, data demand, data supply**

I. Introduction

In our information society, data becomes increasingly a commodity and the basis for many products and services. Examples are Open Data, Linked Data or Big Data applications and services, such as government data portals[1], reviews, feedback, and product suggestion on e-commerce websites, weather emergencies forecast[2], patient monitoring[3], citizen participation and decision-making[4], etc. In recent years, in order to reflect this dataification [1], the concept of *data value chains* was introduced, building upon the concept of traditional value chains for tangible products [2]. The rationale of a data value chain is to extract the highest possible value from data by modifying, processing and re-using it.

The value chain model describes value-adding activities that connect an industry's supply side, such as raw materials and production processes, to its demand side, such as sales and marketing. The value chain model has been used to analyse and assess the linked activities carried out within traditional industries in order to identify where, within these activities, value is created. As successful as the value chain concept was to achieve this aim, during these last years products and services are becoming increasingly digital, and exist in a more non-tangible dimension [3]. In addition, the traditional value chain model does not consider when information is used as a source of value in itself [4]. Thus, the original concept of value chain is becoming an inappropriate method with which to identify value sources in today's industries that produce non-tangible products [3].

In this paper we analyse existing implementations of data value chains and build upon previous definitions with the aim of providing the best characterisation. Using a methodology similar to defining a life-cycle, we hence define the *Data Value Network* (DVN). Our definition caters specifically for non-tangible data products, thus we focus on the aspects specific to data that differ from the definitions of value chains in literature. Based upon our definition, we focus on the creation and consumption of data and propose a *Demand and Supply Distribution Model*. We provide this model as an attempt to give an insight on how an entity can start a DVN on a particular data product. As a concrete implementation, we also provide a prototype of the proposed model through a *Demand and Supply as a Service* application. Through the DVN and the Demand and Supply Distribution Model we propose, we aim to highlight the economic potential of a data product.

II. Background and Related Work

Many works in literature focus on various aspects in the context of value creation, such as how to best exploit value creation to achieve economic benefit [5], discussions on value creation within specific domains, such as mobile commerce [6], open data [7], and e-government [8], or how to achieve competitive advantage [2], [9], [10]. Despite this variation in literature, there are few papers that focus on the actual process of creating value: the value chain.

The term *Value Chain* was first introduced by Porter [2] in 1985. Porter defines a value chain to be the strategically relevant interdependent activities undertaken by a firm in order to achieve its goal. The value chain can be considered as a tool that enables the analysis of the interactions between the different activities in order to identify the sources for competitive advantage. Building upon Porter's definition, Lee and Yang [11] define the *Knowledge Value Chain*; a value chain for knowledge. In [12], Crié and Micheaux provide us with a more generic value chain than Lee and Yang, including raw data in their definition. Peppard and Rylander [3] also discuss a value chain that is more suited

[1]https://open-data.europa.eu/en/data/
[2]http://centrodeoperacoes.rio/
[3]http://www.immunizeindia.org/
[4]https://www.fixmystreet.com/

where the product in question is digitised, and thus non-tangible. In line with more recent popular themes, Miller and Mork [13] and Latif et al. [14] focus on big data and Linked Data respectively. Miller and Mork discuss the data value chain concerning all required actions in aggregating heterogeneous data in an organised manner and creating value (information/knowledge) that can influence decision making. In contrast, Latif et al. propose the *Linked Data Value Chain*. Motivated by the still limited commercial adoption of the Semantic Web, the authors aim to drive the Semantic Web and the use of Linked Data closer to commercial entities.

Nowadays, in a digital data-centric world, the cost of processing data has drastically decreased, and the access to data from multiple sources such as networks, sensors, and the Internet, has skyrocketed the availability of data. Coupled with the dramatic decrease in the cost of data storage, this is enabling huge datasets to be generated or captured, stored, and processed. Alas, value chains such as the one defined by Porter [2] do not cater for data-specific activities, or even for activities where information is used as a source of value in itself [4]. This makes the original value chain model inappropriate to characterise a data value chain, where the data is the product being enriched with the aim of creating value. Furthermore, this enhanced use of data can affect various industries, such as finance, energy, and transport, by increasing their productivity [15], and new innovative uses and services can be developed on the value-added data, impacting the economy and society as a whole.

Compared to the literature we discussed above, in this article we target the domain of *data* value chains. We identified the lack of literature that discusses the creation of value on a data product, as well as the actual processes used to create this value.

III. The Data Value Network

After considering existing value chain definitions, and identifying different real-life data value chains and the contained activities, we define a **Data Value Network** (DVN). Similarly to a life-cycle, the DVN maps the on-going processes through which value is created upon a data product. We included common activities executed on data products, where the final aim is usually the consumption of the data product. Due to the differing order of executing the relevant activities, a star network with the data product as its central node was deemed to be the best way to represent the interactive nature of adding value to data products. We define a *DVN* to be:

> *A set of independent activities having the aim of creating value upon data in order to exploit it as a product*

where different **actors** (e.g. data producers, data consumers) can participate by executing one or more **activities** (e.g. Data Discovery, Data Exploitation), and each activity

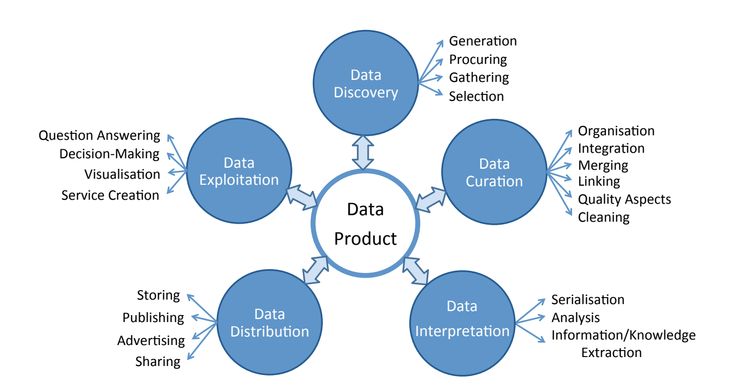


Figure 1. The Data Value Network (Activities and Actions)

can consist of a number of **actions** or value creation techniques, (e.g. Gathering, Visualisation, Service Creation). In turn, each action can consist of one or more **data value chains**, since they might need a series of processes to be executed in order (e.g. visualisation requires identifying the data to visualise, then deciding on a visualisation method, then rendering the visualisation).

The DVN thus has the following features which characterise it and distinguish it from the value chain definitions covered in Section II:

Non-Tangible Data Product: Data can be re-used over and over, even for different purposes than the one it was originally planned for, until the data becomes outdated.

Non-Sequential: The DVN does not necessarily follow a sequential structure, rather, any activity can follow, or precede, any other activity.

Multiple Actors: One or more actors can participate in a DVN to produce value within an activity. Actors can also collaborate in order to co-produce value.

Nested Value Chains: Each activity can be broken down into further, more specialised actions, each of which can be a data value chain within itself.

Recurring Value Network: As opposed to the value chain which ends with the consumption of a product, the DVN can recur as long as the data in question is still relevant.

Independent Activities: Unlike the value chain defined by Porter [2] the value creating processes are not interconnected, and can exist independently.

We portray our definition of the DVN in Figure 1, where the activities all belong to a **data-centric** domain. Here the data product is central to the DVN, as data, in whatever state it is in, can be considered to be a product and consumed. There are five activities, namely *Data Discovery*, *Data Curation*, *Data Interpretation*, *Data Distribution*, and *Data Exploitation*. Each activity is also made up of a number of actions. While not exhaustive the listed actions are, according to existing literature, the most common and generic processes that can be executed on a data product.

IV. Demand and Supply Distribution Model

We here focus on the Data Discovery and Data Distribution activities as core enablers to the DVN to propose a Demand and Supply Distribution Model (Figure 2). These two activities are crucial to the DVN in that without the existence (creation) of data, and its consumption (achievable through data distribution), there can hardly be any value creation upon a data product. In other words, if the data does not exist, or no one knows it exists, no value can be created upon it. The Demand and Supply Distribution Model hence provides an entry point for stakeholders to create value and participate in a DVN by enabling and enhancing the Data Discovery and the Data Distribution activities.

Entities participating as data producers or publishers in the DVN can be overwhelmed by the amount of competition in the global market. Likewise, data consumers can find it difficult to identify whether the data product they need is already on the market. Moreover, if the data is created with a specific use case in mind, it might be difficult to envision or implement it's use in a different domain. This model we propose can be a solution to these problems, where information about data products resulting from entities' DVNs are indexed in a knowledge base, making them available for easier search and discovery. Using this knowledge base, data consumers can easily identify publishers or producers that are providing the data product that they require. Similarly, data producers can be aware of the data products already on the market, thus having the opportunity to target a niche, if it exists, rather than attempting to compete with established data producers. Basically, this model maps the real-life data demand and supply picture. By following this model stakeholders can hence optimise their process of participating in the value creating process upon data products by having a clear picture of the supply and demand, and acting accordingly.

A. *Demand and Supply as a Service*

In order to act as proof of concept, we created a cloud service in the form of a portal. We provide the Demand and Supply as a Service (DSAAS) as an entry point to the DVN and hence also to the *Economic Data Ecosystem* (Online: http://butterbur22.iai.uni-bonn.de/dsaas/). The portal caters for two discrete roles, reflecting the Demand and Supply Distribution Model, namely data producers (Supply) and data consumers (Demand), and aids the value creation process through enabling and enhancing data discovery and reuse, collaborations, and providing contributions to the data market.

The DSAAS provides two different ways for consuming data; a faceted browser and a RESTful API. The faceted browser enables data consumers (humans) to browse the Data Supply and Demand Knowledge Base (as shown in Figure 2) of existing data that they can consume or even contribute to. The RESTful API, on the other hand, provides automated access to the Knowledge Base. This enables third parties to provide their own applications based on the available data. The DSAAS provides the following functions:

- Browsing existing datasets;
- Adding new data;
- Browsing data requests;
- Adding a new data request.

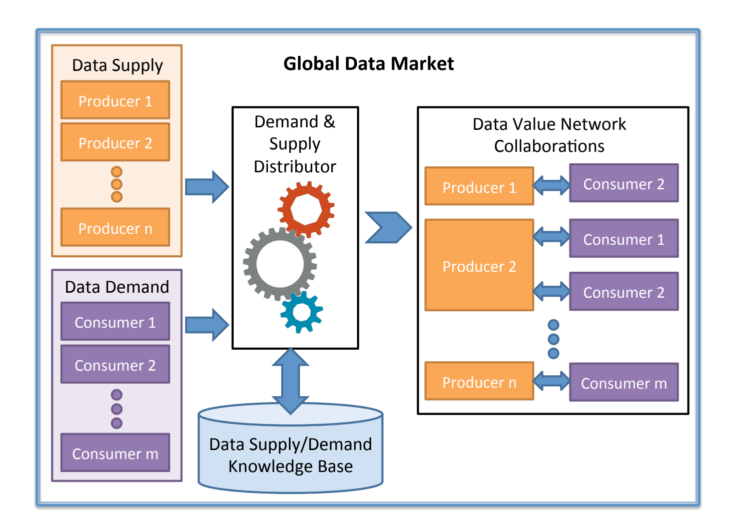


Figure 2. Demand and Supply Distribution Model

In order to best represent the supply and demand of data, we defined the **Demand and Supply Ontology**[5] (DSO) to act as the underlying schema to the DSAAS. The DSO (along with the DSAAS) improves upon existing schemas and initiatives such as DCAT, CKAN[6], and Datahub[7] in that it enables us to represent not only the dataset in question (resulting in a catalogue of datasets), but also all the involved actors, as well as their relationships with the data at hand, hence providing some context on the provenance of the data.

With the current functionality, the DSAAS provides the following benefits:

- Enables quick identification of current demand and supply of datasets, allowing users to target any niche in the data market;
- Eases the discovery of relevant datasets, and hence the participation in a DVN;
- Provides information about different uses of a dataset;
- Enables potential collaboration with other actors;
- Enables users to advertise existing datasets to encourage consumption; and
- Allows users to ask for datasets.

With the aim of asserting the potential benefits of using the DSAAS, we lead out a preliminary evaluation where a number of data producers and consumers were requested to

[5] https://w3id.org/dso
[6] http://ckan.org/
[7] https://datahub.io/

fill out a survey[8]. At the moment of writing, twelve persons responded the survey[9], of which nine are both data consumers and publishers, whilst two are only data consumers. For the questions in the survey which directly concern the foreseen benefits of using the DSAAS, we use the Likert scale to evaluate the degree to which the evaluators agree with the specified benefits. The results of this evaluation indicate that overall the survey responders agree with the benefits of using the DSAAS that we portray. Ten out of twelve respondents agree that the DSAAS encourages dataset sharing and consumption, and eleven respondents agree that it can help the data acquirement process. Whilst there are varying opinions on the benefits, the majority of the respondents always agrees that the tool will improve their participation in the data market. These results, while not conclusive, certainly indicate the potential of our approach.

In order to further establish its validity, the DSAAS is already being used in the ODINE Project[10] [11]. The latter is an open data incubator that provides access to hundreds of companies and SMEs working on open data businesses. Starting in July 2015 (and ending in August 2016), the calls for such entities have already attracted the participation of over 300 companies or SMEs. Use cases and datasets used within the latter SMEs are being fed into the DSAAS knowledge base, hence creating a network of connections and collaborations between the datasets and their producers/consumers. We envisage that once the knowledge base is more substantial, we can also provide additional functions such as a crowdsourced effort for knowledge base curation, importing of existing data catalogs, and the provision of a recommender system built on top of the knowledge base.

V. Conclusion

The increasing dataification within our information society has required the need for the specification and implementation of new value chains. With the aim of projecting our vision of generating a new *Economic Data Ecosystem* based on data value chains, we propose the *Data Value Network* (DVN). The DVN models the co-production of value through the interaction of the involved stakeholders. We also define the *Demand and Supply Distribution Model*, which provides an insight on how an entity can successfully enter the global data market, whilst maintaining a competitive edge. The Demand and Supply as a Service (DSAAS) application acts as concrete implementation of the proposed model. Acting as a dynamic leveller, this service enables stakeholders to more easily advertise existing data products, or otherwise create a request for specific data. This match-making service has the potential of creating a sustainable environment of data re-use, enhancing the value-creation cycle within the DVN.

[8]Complete survey: https://goo.gl/2N3nnZ

[9]Survey responses: https://goo.gl/NzxVI2

[10]https://opendataincubator.eu/

[11]https://www.theguardian.com/technology/2014/nov/04/eu-commits-144m-to-support-open-data-across-europe

References

[1] N. Cukier and V. Mayer-Schoenberger, "The Rise of Big Data," *Foreign Affairs*, 2013. [Online]. Available: https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data

[2] M. E. Porter, *Competitive Advantage: Creating and sustaining superior performance*. NY: Free Press, 1985.

[3] J. Peppard and A. Rylander, "From Value Chain to Value Network:," *European Management Journal*, vol. 24, no. 2-3, Apr. 2006.

[4] J. F. Rayport and J. J. Sviokla, "Exploiting the Virtual Value Chain," *Harvard Business Review*, vol. 73, p. 75, 1995.

[5] R. Amit and C. Zott, "Creating Value Through Business Model Innovation," *MIT Sloan Managment Review*, no. 53310, 2012.

[6] S. J. Barnes, "The mobile commerce value chain: analysis and future developments," *International Journal of Information Management*, vol. 22, no. 2, pp. 91–108, Apr. 2002.

[7] D. Schlagwein and D. Schoder, "The Management of Open Value Creation," in *HICSS-44*. {IEEE} Computer Society, 2011.

[8] A. Wassenaar, "E-Governmental Value Chain Models," in *11th International Workshop on Database and Expert Systems Applications (DEXA'00), 6-8 September 2000, Greenwich, London, {UK}*. {IEEE} Computer Society, 2000.

[9] R. Kaplinsky and M. Morris, "A Handbook for Value Chain Research," no. September, 2002. [Online]. Available: http://oro.open.ac.uk/5861/

[10] M. E. Porter and V. E. Millar, "How information gives you competitive advantage," *Harvard Business Review*, vol. 63, no. 4, pp. 149–162, 1985.

[11] C. C. Lee and J. Yang, "Knowledge value chain," *Journal of Management Development*, vol. 19, no. 9, pp. 783–794, 2000.

[12] D. Crié and A. Micheaux, "From customer data to value: What is lacking in the information chain?" *Journal of Database Marketing & Customer Strategy Management*, vol. 13, no. 4, Jul. 2006.

[13] H. G. Miller and P. Mork, "From Data to Decisions: A Value Chain for Big Data," *IT Professional*, vol. 15, no. 1, Jan. 2013.

[14] A. Latif, A. Us Saeed, P. Hoefler, A. Stocker, and C. Wagner, "The Linked Data Value Chain: A Lightweight Model for Business Engineers," in *Proceedings of International Conference on Semantic Systems*, 2009, pp. 568–576.

[15] O. Tene and J. Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," 2013. [Online]. Available: http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1